Navy Personnel Research and Development Center
San Diego, California 92152-7250
TN-94-6 October 1993

4

An Assessment of the Statistical and Editorial Output of Text Analysis Programs

Paul H. Radtke

NPRDC-TN-94-6 October 1993

An Assessment of the Statistical and Editorial Output of Text Analysis Programs

Paul H. Radtke

Reviewed by
Carol A. Robinson

Approved and released by
J. C. McLachlan
Director, Training Research Department

Navy Personnel Research and Development Center
San Diego, CA 92152-7250

Identification of specific equipment and software is for documentation only, and does not imply endorsement.

Correct Grammar is a trademark of Lifetree Software, Inc.
Editor is a trademark of Modern Language Association.
Grammatik 5 for Windows is a trademark of Reference Software International.
Power Edit is a trademark of Artificial Linguistics.
RightWriter is a trademark of RightSoft, Inc.
Windows is a trademark of Microsoft Corporation.
Writer's Workbench is a trademark of AT&T Bell Laboratories.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
October 1993

3. REPORT TYPE AND DATE COVERED
Final—February-October 1992

4. TITLE AND SUBTITLE
An Assessment of the Statistical and Editorial Output of Text Analysis Programs

5. FUNDING NUMBERS
Program Element: 0602233N
Work Unit: RM33T23.04

6. AUTHOR(S)
Paul H. Radtke

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Navy Personnel Research and Development Center
San Diego, California 92152-7250

8. PERFORMING ORGANIZATION REPORT NUMBER
NPRDC-TN-94-6

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research (Code 461)
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Functional Area: Training
Product Line: Curriculum Acquisition, Development and Revision
Effort: Automated Instructional Development

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This study assessed commercially available text analysis programs for their accuracy and usefulness to novice writers. Two topics were examined: the reliability and accuracy of scores and indicators, and the level and validity of the editorial comments. Six text analysis programs were assessed by comparing their output for two previously published Navy technical publications. The analyzers produced different word and sentence counts, and standard readability scales, suggesting that each uses different algorithms to analyze text material. The analyzers' statistical measures may be sensitive to the complexity and format of text samples, but not to their size. The analyzers varied on the number and level of comments, the validity of the comments, and the types of problems they detect. The text analyzers are able to detect problems of usage, sentence length, and other low-level mistakes. They are less able to detect problems involving the relationship between parts of sentences. Specific problems created by the unique characteristics of technical writing are identified (e.g., technical terminology, passive sentence forms, numbers).

The analyzers assessed in this study are suitable to supplement, but not replace, traditional editing. If accompanied by additional guidance, they might be used as tutorial aids.

14. SUBJECT TERMS
Text analysis, grammar checker, style checker, personal computer software, readability.

15. NUMBER OF PAGES
31

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

Foreword

This research was conducted within Exploratory Development (Program Element 0602233N, Work Unit RM33T23.04) and sponsored by the Office of Naval Research (Code 461).

This report presents an assessment of the statistical and editorial output of six commercially available, text analysis computer programs. The objective of this research was to establish the accuracy, consistency, and validity of these tools, and assess their possible usefulness to Navy technical writers.

J. C. McLACHLAN
Director, Training Research Department

Summary

Problem

The potential use of computer-based text analyzers to aid Navy technical writers has been questioned because of concerns about the adequacy of the tools that are currently available.

Objective

The objective of this study was to assess commercially available, text analysis computer programs for their accuracy and usefulness to novice writers. Two areas were examined: the reliability and accuracy of their scores and indicators, and the level and validity of the editorial comments they generate.

Approach

Six commercially available text analysis programs (i.e., text analyzers) were assessed by comparing and evaluating their statistical output and editorial comments for published text typical of the material produced by Navy technical writers.

Results

The text analyzers produced different word counts, sentence counts, and various standard readability scores, suggesting that each text analyzer uses a different formula and algorithm to analyze input text material. There is some indication that the text analyzers' statistical measures are sensitive to the general complexity and format of text samples, but not to the size of the sample. The text analyzers varied in the number of comments they made, the level of analysis, the validity of the comments, and the specific types of problems they detected in the text.

Conclusion

The text analyzers are able to detect problems of usage, sentence length, and other low level grammatical mistakes. The text analyzers are less able to detect or correctly diagnose problems involving the relationship between parts of sentences or among sets of sentences.

Specific problems are created for computer-based text analyzers by the unique forms and requirements of technical writing. Among the problems identified are the frequent use of uncommon words and phrases; the use of long, complex sentences; the extensive use of passive forms; and the frequent inclusion of numbers in the text.

The text analyzers assessed in this study are suitable to supplement, but not replace, traditional editing, or could be used as tutorial aids for subject matter experts with little writing or editing experience.

Contents

List of Tables

Introduction

Problem

Each year the Navy publishes thousands of technical documents written by subject matter experts who have little experience or training in writing and editing. These documents must pass through an extensive editorial review process to ensure that they meet established standards of format, grammar, usage, and readability. Recently developed computer-based writing aids, such as grammar, style, and spelling checkers, may offer a way to improve the quality of written materials prepared by subject matter experts/writers without requiring additional training in writing, or relying exclusively on professional editors to correct errors. Several concerns, however, have been raised about the use of these automated text analysis tools, also called text analyzers. First, the computer-based tools that are currently available are primarily designed for personal or business use, and may not meet the needs of technical writers. Second, published reviews and informal comments suggest that these tools lack the accuracy and practical usefulness necessary to justify their use in the editorial process (Constanza, 1992; Matzker, 1992; Smith, 1992). Finally, a recent survey found that many writers and editors resist using computer text analyzers, because of a perception that these tools create additional, unnecessary work by reporting an unacceptable number of false "mistakes" (Duffy, in press).

Additional concerns focus on the validity and reliability of the scores and indexes provided by these text analyzers, including several standard scales of readability. Informal evidence suggests that the scores reported by automated text analyzers differ widely, which raises concern about how different text analyzers compute the scores they report. This concern is important if text analyzers are used to verify that documents meet established Department of Defense (DoD) readability criteria, or if they are used in research in which the reading characteristics of sample text is a variable.

Objective

The objective of this study was to assess commercially available, text analysis computer programs for their accuracy and usefulness to novice writers. Two areas were examined: the reliability and accuracy of the scores and indicators, and the level and validity of the editorial comments they generate.

Background

This research continues earlier work on text analyzers conducted for the Navy Personnel Research and Development Center (Duffy, 1991). That work focused on the usability of these tools and informally examined the validity and consistency of the text analyzers' output. Duffy concluded that the capability of current text analysis technology was still relatively low. He also indicated that there were several problems with the amount and quality of the information provided by these analyzers, and how it was presented. First, he noted that the text analyzers tended to produce large numbers of comments, even for previously edited text. He speculated that the amount of information and advice provided by the analyzer might overwhelm writers and discourage its use. Second, he noted the tendency of the analyzers to make many invalid or misleading comments, which might cause writers to discount the usefulness of all comments. Third, he questioned whether writers who required outside help to identify a problem in their

writing would understand the problem well enough to make the necessary changes. Finally, he questioned whether the use of these automated editors would improve the quality of the final text, or the efficiency of the text production process. He noted that the use of text analyzers might slow the production process by adding a redundant search for basic grammar errors that a human editor would repeat in the search for more complex errors.

Usability is an important characteristic of a text analyzer, because it affects the efficiency of the editing process and indirectly influences the willingness of the user to use the program. However, the core value of a text analyzer is in the quality of the information it provides to users. Consequently, the focus of this assessment is on the text analyzers' output—the content, consistency, accuracy, and completeness of the information provided about the text. We did not try to judge the absolute value of the information to the user, but determined only whether the text analyzers reported valid errors and made reasonable critiques of the actual text.

Approach

Text Analysis Programs

The following five text analyzers were assessed in this report:

- *Correct Grammar*, Version 3.0 by Lifetree Software, Inc.
- *Editor*, Version 4.0 by Modern Language Association.
- *Grammatik 5 for Windows* by Reference Software International.
- *Power Edit*, Version 1.0 by Artificial Linguistics.
- *RightWriter*, Version 3.1 by RightSoft, Inc.

Each text analyzer is commercially available and runs on a standard personal computer (PC) running DOS. The *Grammatik 5* text analyzer requires the Windows 3.0 environment. The other text analyzers run under the standard DOS environment. Each text analyzer has a minimal memory and storage space requirement, which is described in Appendix A.

In addition to the five PC-based text analyzers, we examined the output of the *Writer's Workbench* text analyzer, a set of analytic programs developed by AT&T Bell Laboratories, that examines many of the same characteristics of text as the five commercial text analyzers. Although the *Writer's Workbench* text analyzer only runs on larger computers, it provides a benchmark against which to compare the less expensive and more widely-available PC-based text analyzers.

Text Samples

The text samples used in the analysis were taken from two Navy publications:

- NAVSHIPS 93752, *Technical Manual for Radio Transmitting Set AN/WRT-4(XN-2)*, Section 4, Principles of Operation, 4-1 to 4-2c(2)(d) (Naval Sea Systems Command, 1961).

- NAVEDTRA 10185-C1, *Gunner's Mate G 3&2*, Chapter 5, Basic Mechanisms, pp.5-2 to 5-4, and 5-21 to 5-25 (Naval Education and Training Command, 1986).

We used published text rather than draft material to reduce the number of grammatical errors in the text. These particular sources were selected because they are typical of the technical material produced for Navy personnel, but contain distinctive prose and format styles.The first source, NAVSHIPS 93752, is a technical reference manual that presents detailed information about a radio set's design, operational characteristics, installation, operation, and maintenance. Because it is a reference document, its prose style is formal and precise, and it follows a quasi-outline format. In the remainder of this report, we will refer to this text as the *radio sample*.

The second source, NAVEDTRA 10185-C1, is a rate training manual that provides the information enlisted personnel need to learn to pass advancement examinations. The text sample taken from this source describes a variety of mechanical devices found in the equipment used and maintained by Gunner's Mates. Although the content is technical, the prose in this manual is less formal and stylized than that of the reference manual. We will refer to this text as the *mechanism sample* in the rest of this report.

Text samples of about 500, 1,000, 1,500, and 2,000 words were selected from both publications, and converted to standard ASCII text files.The samples preserved the structure of the sources by terminating at the end, rather than within, a paragraph. Thus, the samples contained slightly more words than the target number. Each successively larger sample contained the text of the smaller sample plus additional text. Thus, the smaller text samples were subsets of the 2,000 word sample. When referring to the different samples in terms of their size, we use 500, 1,000, 1,500, and 2,000 although the actual word counts of these samples is slightly higher.

The same text samples were used in each analysis. However, the text analyzer, *Editor*, required a special format for sentences submitted in ASCII form.To function properly, each sentence submitted to this analyzer had to end with a carriage return. Separate files using this format were created from the files used with the other text analyzers. No other changes were made to the text.[1]

Procedure

The text samples were submitted for analysis to the text analyzers according to the instructions provided in the accompanying documentation. Although several of the text analyzers are designed to be used interactively, with output provided directly on the computer screen, the output was converted to paper printout, to provide a permanent record of the comments.

To compare the comments produced by the text analyzers, the two 2,000-word text samples were broken down into passages, consisting of sentences or headings. The radio sample contained 119 passages, and the mechanism sample contained 138 passages.

The analysis of the editorial output was confined to simple number and percentage comparisons. We considered statistical analyses for significance to be impractical because of differences in the information provided by the different text analyzers.

Text analyzers produced two kinds of output: estimates of word count, sentence count, and a variety of other descriptive statistics; and editorial comments that identify possible grammatical errors in the text. The descriptive statistics included various scores of readability.

[1]Despite several edits of the sample text, typographical and transcription errors were found in the samples after they had been submitted to the text analyzers. Thus, the samples were not exact replicas of the source texts and some of the errors reported were not in the original material.

Readability Scores

Readability scores are quantitative measures of the relative difficulty of a piece of text. They are typically calculated by adding a number of weighted variables, such as the average number of letters or syllables per word, or the average number of words per sentence in the text. For example, the formula for computing the Flesch-Kincaid Grade Level score (Klare, 1984) is:

Grade Level = .39 X Average Number of Words per Sentence
+ 11.8 X Average Number of Syllables per Word - 15.59.

The measures are defined as estimates of the reading grade level of the text, the percentage of functionally literate persons able to understand the text, or as a pure numeric score having no performance or behavioral referent. A higher grade level score, or a lower comprehension percentage, indicates a more difficult text. In many instances, the scores are translatable from one form to another. The Flesch Reading Ease measure can be computed as both a grade level estimate and a comprehension percentage, described as the percentage of persons with a fourth grade education who are able to answer three quarters of the test questions asked about the passage (Flesch, 1948). The formulas for computing the readability scores reported by the text analyzers are presented in Table 1.

Table 1

Formulas for Computing Readability Scores and Indexes

Name	Formula
Coleman-Liau	Grade Level = 5.89 X Average Letters per Word - .3 X Average Sentences per 100 Words - 15.8 (Coleman & Liau, 1975).
Flesch Reading Ease	Reading Ease Score = 206.8 - 1.015 X Average Words per Sentence - 84.6 X Average Syllables per Word (Flesch, 1948).
Flesch Grade Level	Grade Level is derived from Flesch Reading Ease Score as follows: If the Reading Ease Score is less than 30, then Grade Level = 17. If the Reading Ease Score is greater than or equal to 30, but less than 50, then Grade Level = 50 - Reading Ease Score / 6.66 + 13. If the Reading Ease Score is greater than or equal to 50, but less than 60, then Grade Level = 60 - Reading Ease Score / 5 + 10. If the Reading Ease Score is greater than or equal to 60, but less than 70, then Grade Level = 70 - Reading Ease Score / 10 + 8. If the Reading Ease Score is greater than or equal to 70, but less than 100, then Grade Level = 100 - Reading Ease Score / 10 + 5. If the Reading Ease Score is greater than or equal to 100, then Grade Level = 4 (formulas adapted from *Writer's Workbench* source code).
Flesch-Kincaid	Grade Level = .39 X Average Words per Sentence + 11.8 X Average Syllables per Word - 15.59 (Klare, 1984).
Gunning Fog	Grade Level = 0.4 X Average Words per Sentence + Total Number of Words with 3 or more syllables (Gunning, 1952).
Kincaid	Grade Level = 11.8 X Average Syllables per Word + .39 X Average Words per Sentence - 15.59 (formula adapted from *Writer's Workbench* source code).
Kincaid Auto	Grade Level = 4.71 X Average Letters per Word + .5 X Average Words per Sentence - 21.43 (formula adapted from *Writer's Workbench* source code).

Results and Discussions

We present the results of the analysis in two sections. The first section discusses the statistical output of the text analyzers, with emphasis on the readability scores. The second section discusses the comments produced by the text analyzers.

Statistical Output

Each text analyzer produces a different set of statistical scores and indicators. Consequently, we are able to compare the analyzers on only a few of the statistics reported: sample word counts, readability scores, and miscellaneous scores and indexes.

Word Count

The most common statistical output provided by the text analyzers is the number of words in the text samples. Since most scores or indexes of readability are based, in part, on this variable, any differences among the text analyzers would result in differences in these indicators. Table 2 presents the word counts for four of the five PC text analyzers and *Writer's Workbench.* One text analyzer, *Power Edit,* does not provide word count.

Table 2

Word Count by Text Sample Size and Text Analyzer

Text Sample/ Manual Count	Word Count[a]									
	Correct Grammar		*Editor*		*Grammatik*		*RightWriter*		*Writer's Workbench*	
N	*N*	*D*	*N*	*D*	*N*	*D*	*N*	*D*	*N*	*D*
Radio Sample										
559	546	-13	561	+2	564	+5	585	+26	559	0
1,024	1,004	-20	1,024	0	1,029	+5	1,065	+41	1,019	-5
1,571	1,547	-24	1,570	-1	1,576	+5	1,630	+59	1,560	-11
2,092	2,058	-34	2,088	-4	2,098	+6	2,176	+84	2,087	-5
Mechanism Sample										
563	555	-8	563	0	555	-8	567	+4	563	0
1,025	1,014	-11	1,025	0	1,014	-11	1,036	+11	1,025	0
1,564	1,553	-11	1,568	+4	1,557	-7	1,588	+24	1,565	+1
2,035	2,024	-11	2,041	+6	2,029	-6	2,078	+43	2,038	+3

Notes. *N* = Number of words, *D* = Difference from manual count.
[a]*Power Edit* does not provide a word count.

Table 2 shows that the text analyzers' word counts differed—sometimes substantially—from the manual word count and from each other. Only two of the text analyzers, *Editor* and *Writer's Workbench*, reported any word counts that were identical to the manual word counts. Of 40 possible word count comparisons, the text analyzers agreed with the manual word count only six times. In only four cases of 80 possible pair-wise agreements did two text analyzers' word counts agree with each other.

There are no agreements between the analyzers' word counts for any of the radio samples. *Correct Grammar* and *Grammatik* produced the same word counts for the 500 and 1,000 word mechanism samples. *Editor* and *Writers' Workbench* also produced identical word counts for the 500 and 1,000 word mechanism samples, but the analyzers each produced different word counts for the 1,500 and 2,000 word mechanism samples.

We cannot identify a single reason for the differences in the word counts. Likely sources of the differences are the headings, labels, and technical nomenclature that include single letters, numbers, and hyphenated labels (e.g., "4-1"). The 2,000-word radio sample and the 2,000-word mechanism sample contain 66 and 44 of these "words," respectively. The text analyzers produced fewer agreements and reported larger differences for the radio sample than for the mechanism sample.

The manual count of the text samples included section headings. Single numbers and letters were counted as single words (e.g., "Figure 4" was counted as two words), although hyphenated labels or names were treated as single words (e.g., "AN/WRT-4(XN-2)" was counted as a single word). A text analyzer that ignored headings or treated single letters and numbers as part of a single word would produce a lower word count than the manual count. A text analyzer that treated hyphenated labels or words as multiple words might produce a higher word count than the manual count. The number of hyphenated words, however, was not sufficient to explain the large differences noted in Table 2, even when words that are normally hyphenated are included (e.g., "up-and-down").

Readability Scores

Table 3 presents the readability scores and other indexes reported by four of the text analyzers. *Editor* provides no statistical estimate of readability. *Power Edit* provides a number of summary scores, collectively called a "Style Portrait," in the form of line figures to which numeric values are difficult to assign. Consequently, the scores are not reported in Table 3. The Style Portrait figures for the two text samples are reproduced in Appendix B.

The scores and indexes reported in Table 3 reflect a difference between the two text sources; that is, the radio sample is a more difficult piece of text than the mechanism sample. The mechanism samples have a reading grade level between grades 7 and 11, whereas the radio samples have a reading grade level between grades 10 and 15. Changes in the other scores and indexes tend to parallel the changes in grade level estimates.

The scores reported by the analyzers changed as the number of words increase, but the changes were small. However, there were substantial score differences among text analyzers for scales with the same label. The Flesch-Kincaid scores for the mechanism samples reported by *Correct Grammar* were consistently higher, but within one grade level of the Flesch-Kincaid scores reported by *Grammatik and RightWriter*. For the radio samples, the Flesch-Kincaid scores reported by both *Correct Grammar* and *RightWriter* decreased about a grade and a half as the sample size increased from 500 to 2,000 words. However, the scores reported by *Grammatik* were consistently higher than the scores reported by *Correct Grammar* and *RightWriter*, and remained at grade 13 for all but the 1,500 word sample, which increased to grade 14.

Table 3

Readability Scores and Indexes by Text Sample Size and Text Analyzer

Text Analyzer/Readability Scores or Index[a]	Radio Sample 500	1,000	1,500	2,000	Mechanism Sample 500	1,000	1,500	2,000
Correct Grammar								
Flesch-Kincaid	12.0	11.4	11.1	10.5	7.7	8.5	8.3	8.4
Flesch Reading Ease[b]	36.9	40.9	45.3	48.1	64.8	61.1	62.6	62.3
Flesch Grade Level	14	13	12	12	8	8	8	8
Gunning Fog	8.9	8.9	9.0	8.4	7.3	7.8	7.6	7.8
Grammatik								
Flesch-Kincaid	13	13	14	13	8	9	9	9
Flesch Reading Ease	30	26	26	26	65	62	61	62
Gunning Fog	19	18	17	15	10	11	11	11
RightWriter								
Flesch-Kincaid	11.5	11.0	10.3	10.0	7.3	8.1	7.8	7.9
Flesch Reading Ease	41.9	45.1	50.7	52.4	68.3	64.9	66.5	66.5
Gunning Fog	15.7	14.6	13.4	12.9	9.1	10.0	10.0	10.1
Writer's Workbench								
Coleman-Liau	13.8	13.3	12.6	12.3	9.9	10.2	10.1	10.0
Flesch Grade Level	15.9	15.7	14.6	14.3	8.5	8.7	8.5	8.4
Flesch Reading Ease	30.4	31.9	39.2	41.6	65.2	62.6	65.5	65.6
Kincaid Grade Level	13.7	13.5	12.5	12.1	7.8	8.3	7.9	8.0
Kincaid Auto	12.8	12.5	12.0	11.7	8.0	8.5	8.4	8.5

Notes. 1. *Editor* did not provide readability statistics.
2. See Appendix B for *Power Edit* Style Profile.

[a]See Table 1 for formulas used to compute readability scores and indexes.

[b]Reading Ease measure is the percentage of persons with a fourth grade education who are able to answer three quarters of the test questions asked about the passage (Flesch, 1948).

The Flesch Reading Ease and Gunning Fog scores reported by the different text analyzers also appear to have been computed with different formulas, weights, or intervals. While the scores reflect a general difference between the two text samples, the relative size of the score differences between the mechanism and radio samples appears to vary widely among the analyzers. Also, changes between scores as the size of the sample changed vary in both size and direction. These difference in readability scores may be due to the differences in the way words are counted by the text analyzers, as noted earlier. However, they may be due to the characteristics of the two samples. There appears to be greater variability among the scores for the radio samples than among the scores for the mechanism samples. Whatever the validity of the readability scores within a particular text analyzer, they have little general reliability across text analyzers.

Miscellaneous Statistics

Table 4 presents a summary of various miscellaneous statistics generated by the text analyzers. Five text analyzers report the number of sentences in the text; four text analyzers report the average number of words per sentence, and three report the average number of letters and the average number of syllables per word. Two text analyzers report the number of paragraphs and the average number of sentences per paragraph. As with the word counts, the statistics vary among the text analyzers. For the most frequently reported statistic, number of sentences, the text analyzers differ from the manual count in 36 of 40 cases and among themselves in all but six cases.[2]

Table 4

Miscellaneous Statistics by Text Sample Size and Text Analyzer

	Radio Sample				Mechanism Sample			
Text Analyzer/Statistic	500	1,000	1,500	2,000	500	1,000	1,500	2,000
Correct Grammar								
Number of Sentences	34	63	91	129	38	65	100	126
Average Words/Sentence	16.0	15.9	17.0	15.9	14.6	15.6	15.5	16.0
Average Letters/Word	5.1	4.9	4.8	4.7	4.6	4.6	4.6	4.6
Number of Paragraphs	9	17	26	34	6	12	22	28
Average Sentences/Paragraph	3.7	3.7	3.5	3.7	6.3	5.4	4.5	4.5
Average Syllables/Word	1.81	1.77	1.70	1.68	1.50	1.53	1.51	1.51
Passive Sentences	29%	34%	41%	37%	23%	32%	33%	34%
Long Sentences	11%	9%	7%	6%	0%	0%	1%	1%
Comprehension	33	33	43	54	88	88	88	88
Editor								
Number of Sentences	33	61	91	125	42	73	110	138
Number of "To Be" Verb Forms	22	47	73	93	28	47	71	95
Number of Logic Markers	1	2	2	2	5	10	13	16
Number "This" or "It"	3	4	9	11	7	14	23	32
Grammatik								
Number of Sentences	40	73	102	141	38	66	104	132
Average Words/Sentence	14.1	14.0	15.4	14.8	14.6	15.3	14.9	15.3
Average Letters/Word	5.31	5.21	5.08	5.05	4.68	4.71	4.71	4.68
Number of Paragraphs	10	18	27	34	6	13	22	29
Average Sentences/Paragraph	4.0	4.0	3.7	4.1	6.3	5.0	4.7	4.5
Average Syllables/Word	1.92	1.97	1.95	1.96	1.50	1.53	1.54	1.53
Number of Sentences < 12 Words	20	35	43	65	23	32	46	55
Number of Sentences > 40 Words	0	1	1	2	0	0	0	0

[2]The manual counts produced the following estimates of the number of sentences: Radio Samples—500 words, 33 sentences; 1,000 words, 61 sentences; 1,500 words, 90 sentences; 2,000 words, 123 sentences. Mechanism Samples—500 words, 42 sentences; 1,000 words, 73 sentences; 1,500 words, 109 sentences; 2,000 words, 137 sentences. This count includes headings.

Table 4 (Continued)

Text Analyzer/Statistic	Radio Sample 500	1,000	1,500	2,000	Mechanism Sample 500	1,000	1,500	2,000
RightWriter								
Number of Sentences	32	60	93	125	38	65	99	125
Average Words/Sentence	16.6	16.4	16.5	16.4	14.5	15.5	15.5	16.0
Average Syllables/Word	1.75	1.72	1.65	1.63	1.46	1.49	1.47	1.47
Number of Syllables	1,024	1,826	2,686	3,543	830	1,546	2,339	3,033
Number of Words in Sentences	530	985	1,534	2,053	551	1,007	1,536	2,000
Number Unique Words	263	397	484	587	240	366	524	640
Number "Shall"	0	0	0	0	0	0	0	0
Number "Must"	0	0	0	0	1	1	1	1
Number "Can"	1	1	1	1	4	8	9	13
Number "May"	0	0	1	1	0	0	0	3
Strength	.59	.62	.65	.65	.77	.71	.74	.75
Descriptive	.32	.32	.33	.33	.46	.46	.45	.46
Jargon	.00	.00	.00	.00	.00	.00	.00	.00
Writer's Workbench								
Number of Sentences	30	54	83	112	39	67	104	131
Average Words/Sentence	18.6	18.9	18.8	18.6	14.4	15.3	15.0	15.6
Average Letters/Word	5.30	5.21	5.09	5.05	4.72	4.74	4.73	4.70
Number of Content Words	373	670	997	1,350	323	598	917	1,189
Average Letters/Content Word	6.57	6.51	6.45	6.33	6.02	6.06	6.04	6.01
Number of Sentences < 14 Words	10	17	25	35	8	13	21	30
Number of Sentences > 29 Words	3	6	9	12	4	7	8	9
Number of Simple Sentences	19	36	55	77	29	45	71	84
Number of Compound Sentences	4	8	11	12	1	5	7	7
Number of Complex Sentences	5	8	15	18	8	15	22	33
Number of Compound/Complex Sentences	2	2	2	5	1	2	4	7
"To Be" Verbs	46%	53%	54%	53%	57%	48%	46%	45%
Infinitive Verbs	11%	9%	12%	12%	8%	15%	18%	19%
Auxiliary Verbs	7%	3%	4%	3%	14%	15%	12%	14%
Passive Verbs	34%	45%	49%	46%	27%	28%	31%	32%
Nominalizations	3%	2%	2%	2%	4%	4%	4%	4%

Additional evidence that the text analyzers use different rules and formulas to analyze text is the finding that, despite similar estimates of the number of sentences in the mechanism samples, *Grammatik* reported the number of sentences with fewer than 12 words to be consistently higher than number of sentences with less than 14 words reported by *Writer's Workbench*. The statistics also reinforce the conclusion that the radio sample is a more difficult piece of text than the mechanism sample.

Editorial Comments

The statistical information provided by the text analyzers is secondary to the editorial comments they provide on the content, structure, and grammatical correctness of the text. We divided the comments into levels according to the portion of the text they attempt to analyze. The comments were divided among those that refer to (1) individual words, (2) groups of words or phrases within a sentence, or (3) whole sentences or groups of sentences. In the rest of this report, comments will be described as being at the *word, phrase*, or *sentence* level, according to this definition.

Comments at the word level include references to the correct form of abbreviations or the need to avoid specific words because they are vague, pompous, out-of-date, overused "buzzwords," homonyms, words that are often used incorrectly (e.g., "which" vs. "that"), or words with several uses or meanings.

Comments at the phrase level include warnings against the use of overused and wordy expressions, noun strings, or redundant constructions. This level also includes analysis of relationships between words within a sentence, such as using "a" or "an" with nouns beginning with vowels or consonants. Other problems at this level include article-noun, adjective-noun, and verb-object disagreements; the misuse of prepositions; and the overuse of passive verb forms.

Comments at the sentence level include identifying subject-verb disagreements, excessive sentence length, run-on and incomplete sentences, and the incorrect use of conjunctions and semicolons in complex sentences. Comments at this level also warn against overusing certain expressions or constructions across a set of sentences.

It is sometimes difficult to categorize a comment because of the way particular errors are identified by the text analyzers. For example, *Power Edit* identifies every use of a passive verb form throughout the text, apparently following a judgment that *any* use of a passive verb form should be reconsidered by the writer. However, *Grammatik* comments whenever it encounters two passive sentences in a series of ten sentences. This text analyzer apparently follows a rule to avoid the *overuse* of the passive verb form. Under our categorization scheme the *Power Edit* comment is at the phrase level because it focuses on a portion of a single sentence, while the *Grammatik* comment is at the sentence level because it considers how often the passive form is used over a series of sentences. Yet, both text analyzers focus on the same grammatical issue.

This difficulty reflects a basic difference in the text analyzers' approach to detecting, analyzing, and reporting problems in a body of text. Some text analyzers comment on a word or phrase (e.g., "effect" or "affect") whenever it appears, without attempting to determine if the word or phrase is being used correctly. Other text analyzers take the next step and attempt to analyze the word or phrase in the context of the sentence. The first approach virtually guarantees that any misuse of a word or phrase will be noted, but at the cost of an increase in the number of unnecessary comments. The second approach produces fewer, more accurate comments, but increases the probability that the text analyzer will fail to detect a problem or will report a "problem" that does not exist.

Word, Phrase, and Sentence Level

Table 5 presents the number and percentage of editorial comments by each text analyzer for the two 2,000 word samples, and a combined tally of both samples. The comments are broken down by word, phrase, and sentence level.

Table 5

Number and Percentage of Editorial Comments by Text Sample, Comment Level, and Text Analyzer

Text Sample/ Comment Level	*Correct Grammar* N	*Correct Grammar* %	*Editor* N	*Editor* %	*Grammatik* N	*Grammatik* %	*Power Edit* N	*Power Edit* %	*RightWriter* N	*RightWriter* %	*Writer's Workbench* N	*Writer's Workbench* %
Radio Sample (2,000 Words)												
Word Level	2	5.9	54	54.0	33	22.9	9	9.5	4	7.8	2	18.2
Phrase Level	25	73.5	46	46.0	42	29.2	75	78.9	14	27.5	9	81.8
Sentence Level	7	20.6	0	0.0	69	47.9	11	11.6	33	64.7	0	0.0
Total	34	100.0	100	100.0	144	100.0	95	100.0	51	100.0	11	100.0
Mechanism Sample (2,000 Words)												
Word Level	6	26.1	47	58.8	18	19.0	26	17.1	3	9.4	12	60.0
Phrase Level	10	43.4	33	41.2	35	36.8	100	65.8	12	37.5	8	40.0
Sentence Level	7	30.5	0	00	42	44.2	26	17.1	17	53.1	0	0.0
Total	23	100.0	80	100.0	95	100.0	152	100.0	32	100.0	20	100.0
Combined Samples												
Word Level	8	14.0	101	56.1	51	21.4	35	14.1	7	8.4	14	45.2
Phrase Level	35	61.4	79	43.9	77	32.2	175	70.9	26	31.4	17	54.8
Sentence Level	14	24.6	0	0.0	111	46.4	37	15.0	50	60.2	0	0.0
Total	57	100.0	180	100.0	239	100.0	247	100.0	83	100.0	31	100.0

Given identical texts, the text analyzers found different numbers and types of possible grammar and usage problems. *Power Edit* produced 152 comments about the 2,000 word mechanism sample and 95 comments about the radio sample. By contrast, *Correct Grammar* produced only 23 and 34 comments, respectively. Four of the six text analyzers made more comments about the more difficult radio sample than the mechanism sample, but the overall difference in the number of comments made by all six text analyzers on the two samples is small.

The text analyzers also differ in the level of the comments they made about the texts. *Correct Grammar* and *Power Edit* made most of their comments at the phrase level. *Editor* made most of its comments at the word level and made no comments at the sentence level. *Grammatik* and *RightWriter* made most of their comments at the sentence level. The *Writer's Workbench* made comments only at the word and phrase level.

The number and distribution of comments suggest that the text analyzers differ in how they critique textual material and that the differences were consistent for both text samples.

The text analyzers, as a group, commented at least once about most of the passages in the two texts. Of the 119 passages in the radio sample (including headings), only 12 were not critiqued by at least one text analyzer. Of the 138 passages in the mechanism samples, only 23 were not commented at least once. The largest number of comments for any one passage was 17, for a 36-word sentence in the radio sample. The largest number of comments in the mechanism sample was 10 for a 21-word sentence. The average number of comments per passage was 3.7 for the radio sample, and 2.9 for the mechanism sample.

Despite the frequency of comments, there was limited agreement among the text analyzers regarding specific problems in the passages. Of the 435 comments made about the radio sample, only 16 were made by a single analyzer without the agreement of at least one other program. Only 23 of the 402 comments made about the mechanism sample were made by only one text analyzer. However, in no instance did all six text analyzers identify the same problem in the same passage. Further, there were only five passages in the radio sample and only four passages in the mechanism sample in which five text analyzers identified the same problem. Thus, the text analyzers tended to identify different problems within the same passage, usually in agreement with one other analyzer.

The text analyzers agreed most often about relatively simple grammatical problems. For the radio sample, there were 10 passages in which three or more analyzers identified the same problem. Of these, six comments identified a wordy phrase (e.g., "to be referred to as"), two comments identified a sentence as being too long, and two comments identified a problem with the use of multiple semicolons in a sentence. In the mechanism sample, there were 12 instances when three or more analyzers commented about the same problem: Six comments related to the use of excess words, three comments referred to the need to capitalize the first word of a sentence, two comments identified a possible misuse of "which" or "that," and one comment identified a possible conflict between the subject and the verb of a sentence. This last comment is the only instance of three text analyzers diagnosing a problem regarding the relationship between parts of a sentence, rather than flagging a possible misuse of a word or phrase.

Validity

Table 6 presents the distribution of text analyzer comments judged to be valid or invalid by comment level. The decision to categorize a comment as valid or invalid was based on conservative criteria; that is, we labeled a comment as invalid only when it provided incorrect or misleading information. We did not question the validity of comments regarding style, wordiness, or correct usage with which a writer might agree or disagree. We focused only on those comments that required the writer to do additional, unnecessary work to identify and reject the comment.

The most frequent types of invalid comments included misidentifying the function of a word in a sentence (e.g., mistaking a noun for a verb), and misinterpreting the relationship between words (e.g., assuming an adjective-noun relationship between unrelated words). The next most common source of invalid comments was the inability of an analyzer to identify headings, labels, and parenthetical expressions. This resulted in invalid comments regarding sentence completeness, sentence length, punctuation, and article-noun disagreement. The analyzers produced a small number of invalid comments that could not be attributed to a specific interpretive error.

The text analyzers varied widely in the proportion of comments we judged to be valid and invalid. *Correct Grammar* had the highest percentage of invalid comments across the two text samples (54.4%), followed, in order, by *Grammatik* (43.9%), *Editor* (18.9%), *RightWriter* (12.0%), *Power Edit* (10.1%), and *Writer's Workbench* (0.0%). *Writer's Workbench* produced no invalid comments because it focused exclusively on usage problems, which were accepted as a class when we judged a comment's validity.

Table 6

Validity of Editorial Comments by Text Sample, Comment Level, and Text Analyzer

Text Sample/ Comment Level	*Correct Grammar*		*Editor*		*Grammatik*		*Power Edit*		*RightWriter*		*Writer's Workbench*	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Radio Sample												
All Comments												
Valid	16	47.1	74	74.0	72	50.0	90	94.8	45	88.1	11	100.0
Invalid	18	52.9	26	26.0	72	50.0	5	5.2	6	11.9	0	0.0
Word Level												
Valid	2	100.0	32	59.3	14	42.4	9	100.0	4	100.0	2	100.0
Invalid	0	0.0	22	40.7	19	57.6	0	0.0	0	0.0	0	0.0
Phrase Level												
Valid	12	48.0	42	91.3	22	52.4	75	100.0	13	92.9	9	100.0
Invalid	13	52.0	4	8.7	20	47.6	0	0.0	1	7.1	0	0.0
Sentence Level												
Valid	2	28.6	--	--	36	52.2	6	54.5	28	84.8	--	--
Invalid	5	71.4	--	--	33	47.8	5	45.5	5	15.2	--	--
Mechanism Sample												
All Comments												
Valid	10	43.5	72	90.0	62	65.3	132	86.8	28	87.5	20	100.0
Invalid	13	56.5	8	10.0	33	34.7	20	13.2	4	12.5	0	0.0
Word Level												
Valid	6	100.0	41	87.2	17	94.4	23	88.5	2	66.7	12	100.0
Invalid	0	0.0	6	12.8	1	5.6	3	11.5	1	33.3	0	0.0
Phrase Level												
Valid	3	30.0	31	93.9	15	42.9	87	87.0	11	91.7	8	100.0
Invalid	7	70.0	2	6.1	20	57.1	13	13.0	1	8.3	0	0.0
Sentence Level												
Valid	1	14.3	--	--	30	71.4	22	84.6	15	88.2	--	--
Invalid	6	85.7	--	--	12	28.6	4	15.4	2	11.8	--	--
Combined Samples												
All Comments												
Valid	26	45.6	146	81.1	134	56.1	223	89.9	73	88.0	31	100.0
Invalid	31	54.4	34	18.9	105	43.9	25	10.1	10	12.0	0	0.0
Word Level												
Valid	8	100.0	73	72.3	31	60.8	32	91.4	6	85.7	14	100.0
Invalid	0	0.0	28	27.7	20	39.2	3	8.6	1	14.3	0	0.0
Phrase Level												
Valid	15	42.9	73	92.4	37	48.1	163	92.6	24	92.3	17	100.0
Invalid	20	57.1	6	7.6	40	51.9	13	7.4	2	7.7	0	0.0
Sentence Level												
Valid	3	21.4	--	--	66	59.5	28	75.7	43	86.0	--	--
Invalid	11	78.6	--	--	45	40.5	9	24.3	7	14.0	--	--

The text analyzers, as a group, produced a higher percentage of valid comments for the easier mechanism sample (80.6%) than for the more difficult radio sample (70.6%). As a group, the text analyzers were more likely to make an invalid comment at the sentence level (34.0%) than at the word (24.1%) or phrase level (19.8%). The text analyzers reported substantially more invalid comments for the radio sample than the mechanism sample at the word level (39.4 to 9.8%) and the sentence level (40.0 to 26.1%), but there was little difference at the phrase level (18.0 to 21.7%). However, these patterns of valid and invalid comments are inconsistent across the analyzers because of differences in the level of comments they made, and their susceptibility to making errors about certain grammatical problems. These variations among the text analyzers are difficult to summarize in purely quantitative terms. We summarize the general nature of the comments made by the text analyzers, and how these tendencies influenced the comparisons shown in Table 6.

Correct Grammar. *Correct Grammar* reported few comments and a high percentage of those were judged invalid. It made few comments at either the word or the sentence level. The few word level comments were consistently valid, but the sentence level comments were frequently invalid. These patterns were consistent across both text samples. At the phrase level, slightly less than half the comments were valid for the radio sample, whereas only 30% (3 of 10) of its phrase-level comments were valid for the mechanism sample. *Correct Grammar* identified valid problems related to punctuation, sentence length, capitalization errors, prepositional usage, excess words, and "which/that" usage. It produced invalid errors in the diagnosis of run-on and incomplete sentences, verb-object and subject-verb disagreements, and identifying possessive relationships.

Editor. *Editor* produced many comments, all at the word or phrase level, and all but a handful relating to usage problems: commonly misused words and phrases, jargon, wordy expressions, and homonyms. In these areas, the comments were consistently valid. It also reported valid problems with punctuation. However, comments identifying problems with spacing and capitalization, article-noun disagreement, possessive constructions, and the use of archaic terms were frequently invalid. The majority of invalid comments were the result of applying a blanket rule that all whole numbers, such as "3" or "1,200," should be spelled out rather than presented in numeric format. Because of the technical nature of the two samples, such whole numbers were frequently used as labels (e.g., "Table 3") or as constant values (e.g., "1200 kilowatts") and, therefore, this insistence was inappropriate.

Grammatik. *Grammatik* produced a large number of comments, just over half of which were valid. There is no clear pattern to the comments' validity and their level of analysis. *Grammatik* commented most often at the sentence level, and least often at the word level. It reported a higher percentage of valid comments for the mechanism sample (65.3%) than the radio sample (50.0%). It was more successful at diagnosing problems at the word and sentence level, and less successful at producing valid comments at the phrase level. Specific comments successfully identified problems with the overuse of passive verbs, word and phrase usage, spacing errors, abbreviations, sentence length, paragraph length, "which/that" usage, prepositional usage, and the repetitious use of words and phrases over a series of sentences (e.g., using "The" to begin a sentence). Comments identifying problems with verb-object disagreements, possessive constructions, incomplete noun phrases, punctuation, capitalization, and identifying the parts of sentences were inconsistent. It had consistent difficulty identifying incomplete sentences, article-noun disagreements, subject-verb disagreements, and the disagreement of adjectives and nouns. An example of this error was the

analyzer's interpretation of a passage containing the phrase: 3" gearing. The analyzer incorrectly assumed that "3" was a plural adjective modifying the singular noun "gearing." Like *Editor*, *Grammatik* also insisted that all whole numbers should be spelled out.

Power Edit. Of the five PC-based text analyzers, *Power Edit* produced the largest number of comments and had the smallest percentage of invalid comments. About 70% of its comments were at the phrase level, and about 15% each at the word and sentence level. Over 90% of the comments at the word and phrase level were valid, as were about 75% of the comments at the sentence level. *Power Edit* reported a higher percentage of valid comments for the radio sample (94.8%) than the mechanism sample (86.8%). It successfully identified problems of word and phrase usage, punctuation, overuse of passive verbs, noun strings, sentence length, and missing words. It was able to identify problems with compound sentences and to distinguish between sentences and headings better than the other analyzers. However, like the other text analyzers, it was inconsistent in identifying article-noun and subject-verb disagreements. It also had difficulty in identifying incomplete sentences, adjective-noun disagreements, and the main idea of a sentence. Several invalid comments were made because *Power Edit* ignored text enclosed in parentheses.

RightWriter. *RightWriter* produced fewer comments than *Editor*, *Grammatik*, and *Power Edit*, but the percentage of valid comments was high. *RightWriter* made most of its comments at the sentence level and few at the word level. However, it limited the problems it attempted to identify to sentence length, word, and phrase usage problems. It was less successful identifying valid problems with article-noun disagreement, incomplete sentences, and the use of archaic language. *RightWriter* sometimes treated two short sentences as a single sentence, resulting in invalid comments regarding excessive sentence length.

Writer's Workbench. *Writer's Workbench* was the least error prone of the text analyzers because the type of grammatical error it attempted to analyze, word- and phrase-level usage, was always judged valid under our criteria.

Conclusions

The statistical and editorial output produced by these six text analyzers reflect a variety of approaches to the common tasks of analyzing text and reporting problems. This variability in approach is suggested by the lack of agreement among the text analyzers in the kinds of grammatical problems they detected and in the different statistical outputs they produced. These text analyzers, as a group, are also limited in their ability to analyze grammatical problems. At present, they appear to be most effective in identifying word- and phrase-level problems involving word usage, passive verb forms, and wordiness. Their ability to identify more complex problems involving the relationship between parts of a sentence and the diagnosis of whole sentences or several sentences is still primitive and inconsistent. The text analyzers that tried to be helpful to the writer by trying to diagnose more complex problems were more likely to produce invalid comments. While these particular text samples may have biased the outcome for or against specific text analyzers because of the way they approach the task, the broader conclusion that the analyzers are inconsistent and incomplete in their analysis of grammatical problems is well-founded by these results.

Application of Text Analyzers to Technical Material

Throughout the discussion of the results we commented on specific problems encountered by the text analyzers because of the unusual or uncommon structures and usages found in technical writing. In the remainder of this discussion, we summarize and discuss the issues related to the use of these text analyzers with technical materials that are suggested by this assessment.

The technical material reviewed in this analysis posed relatively few unique problems for the text analyzers, despite the peculiar usages, vocabulary, and formats. Some of the difficulties noted in the findings can be reduced by customizing these tools to different text styles and formats, by adjusting the tolerances on certain rules, and by allowing the writer to introduce specialized vocabulary. Since this analysis was run only once on these text samples, the text analyzers frequently flagged uncommon technical terms (e.g., trunnion, zerk, diode) that would be added to the analyzers' dictionary and ignored in subsequent analyses. This feature is useful when such words are used frequently, but is less useful for specific component labels, such as alphanumeric names for electronic components (e.g., "diode A346B"), call letters, and specific kinds of equipment (e.g., "Radio Set AN/GRC-106"). There are too many of these items, and they are used too infrequently to make the introduction of these terms into the text analyzers' dictionary either practical or useful.

A second source of invalid comments is the frequent use of numbers in technical writing. The rule that whole numbers should be spelled out rather than presented in numeric form produced invalid comments for virtually every table reference, component label, heading, and numeric value. The number of these comments is likely to distract the writer, although very little thought is needed to reject or accept the comment when it is encountered.

The text analyzers are capable of identifying and sensitizing technical writers to problems commonly associated with technical writing, such the overuse of passive forms and of long, complex sentences. The use of lists, repetitive formats, and passive sentence forms is common in the technical literature, making these "errors" more prominent than they might be in less formal writing. The analyzers that examined the length and complexity of sentences, and how often certain words or phrases were used produced many "valid" but inappropriate comments that suggested the need to simplify, shorten, or vary the text. The ability to adjust the text analyzer to tolerate these formats would be a useful feature of a text analyzer used to examine technical material.

A large number of invalid comments were due to the text analyzers' misinterpreting the function of words that are used in unconventional ways in technical writing. For example, words, such as relay, switch, and load, that are used as verbs in most normal usage, but are also used as nouns in technical writing, appeared to confuse the text analyzers, leading to invalid comments. However, the inability of text analyzers to identify the parts of a sentence is a potential problem for all forms of writing, technical or otherwise.

Given these limitations, the next question is whether these tools provide enough benefit for technical writers to justify their use. Clearly, the programs reviewed here could not substitute for the review of a human editor. However, they might help subject matter experts/writers improve their early drafts before submitting them to the editor. In addition, the process of submitting a draft to the text analyzer might sensitize writers to certain grammatical issues and encourage them to improve their writing.

If the information provided by a text analyzer is used primarily to instruct writers rather than improve the quality of a specific piece of text, the value of the information may be determined more by

the willingness and ability of writers to understand, evaluate, and apply the information than by the absolute accuracy of the information itself. However, if writers cannot rely on a text analyzer to identify all problems and if the information provided by the program is sometimes unnecessary, misleading, or erroneous, the burden on novice writers is likely to be substantial. How much information would writers want from the analyzer? How many invalid or misleading comments will writers tolerate? How would writers use the information provided? How often do they accept or reject the comments and advice? How do they interpret and use the readability and other quantitative measures provided by these programs? The questions relating to the practical use of the information provided by a text analyzer go beyond the scope of this research and should be explored through the study of actual writers using these text analysis tools over an extended period.

Based on these findings, Navy organizations considering the use of text analyzer programs in their editorial process should also consider how to integrate these tools into the process so as to compensate for their current limitations. The most important implication of this study is that these tools are not an adequate substitute for a human editor, even for relatively mechanical editing tasks. In addition, using these tools to verify that written materials meet current DoD readability standards should be discouraged because of the variability of the scores produced by different analyzers. Other uses of these tools, such as their use to improve the initial drafts of subject matter experts with limited writing experience or as a tutorial device, should be accompanied by warnings against overreliance on the comments. Novice writers should also have access to additional guidance to help them interpret the output.

References

Bureau of Ships. (')61). *Technical manual for Radio Transmitting Set AN/WRT-4(XN-2)* (NAVSHIPS 93752). Washington, DC: Author.

Coleman, M., & Liau T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology, 60* (2), 283-284.

Constanza, W. V. (1992, August). Educational technology product review. *Educational Technology*, 56-57.

Duffy, T. M. (In press). *Designing tools to aid technical editors: A needs analysis*. San Diego, CA: Navy Personnel Research and Development Center.

Duffy, T. M. (1991). *Text analyzers: A review and informal evaluation*. Unpublished research report, San Diego, CA: Navy Personnel Research and Development Center.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221-233.

Gunning, R. (1952). *The technique of clear writing*. New York: McGraw-Hill.

Klare, G. R. (1984). Readability. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 681-744). New York: Longman.

Matzker, J. (1992). Windows brings added dimensions to *The Writer's Toolkit. PC Magazine, 11* (3), 54.

Naval Education and Training Program Management Support Activity. (1986, November). *Gunner's Mate G 3&2* (NAVEDTRA 10185-C1). Pensacola, FL: Author.

Smith. J. (1992, October). Mark your words with grammar checker software. *PC Computing*, 243-252.

Appendix A

Text Analyzer Memory and Storage Requirements

Text Analyzer Memory and Storage Requirements

Correct Grammar, Version 3.0

RAM Requirement:	512K
Hard Drive Space Required:	1MB

Editor, Version 4.0

RAM Requirement:	256K
Hard Drive Space Required:	400K

Grammatik 5 for Windows

RAM Requirement:	2MB
Hard Drive Space Required:	1.6MB and extended or expanded memory if available.

Power Edit, Version 1.0

RAM Requirement:	470 and 1MB extended or expanded memory
Hard Drive Space Required:	12MB

RightWriter, Version 3.1

RAM Requirement:	512K
Hard Drive Space Required:	.5MB

Appendix B

Power Edit: Style Portraits

RADIO SAMPLE

STYLE PORTRAIT

```
                  Sentence            Document
               1  2  3  4  5        1  2  3  4  5
Retention            X ------             X ------
Verb Strength    -X----------        -X----------
Formal        ----X-------         ------------
Verb Clarity   X     ---------      X   ---------
Action        ---------------      ---------------
Unclear       ------               ------
Sensitivity   ------               ------
Climactic     ----X----            ---------
Theme Level       X    ------          X    ------
Incorrect         X
Filler        ---                  ---
Slang
Reflection    ------               ------
Complexity    ----X-               -X----
Ambiguity     -X-                  ---
Visual Form   ---------            ---------
Literary      ---                  ---
Derogatory
Breezy
Abstract      -X----               -X----
Difficulty    ---                  ---
Monotony      ------ X             ----X-
```

----- preferred setting area
X actual setting
'Sentence' shows the last sentence's statistics.
'Document' shows the cumulative document average.

Style	Base	Explanation
Retention	5	How well your reader will remember what was written.
Verb Strgth	5	Measures your choice of good, strong verbs.
Formal	1	Measures how formal your writing is.
Verb Clarity	5	Measures how well actions are being expressed.
Action	5	Measures the verb strength of the main clause.
Unclear	1	Measures the use of unclear words or phrases.
Sensitivity	1	Measures the potential level of mis-communication.
Climactic	1	Is your most important information last?
Theme Level	5	Measures the use of strong topics, actions, focuses.
Incorrect	1	The number of incorrect words or phrases used.
Filler	1	Measures the number of times filler words are used.
Slang	1	Measures the amount of slang words in a sentence.
Reflection	1	Measures the level of author intrusion.
Complexity	1	Measures the complexity of the sentences.
Ambiguity	1	Measures the potential ambiguity in the document.
Visual Form	1	Abbreviations, foreign expressions...
Literary	1	Measures the use of literary words and phrases.
Derogatory	1	Measures the use of derogatory words and phrases.
Breezy	1	Measures the use of folksy and informal expressions.
Abstract	1	Measures the use of abstract nouns in the document.
Difficulty	1	Measures the use of jargon, legalese, code words...
Monotony	1	Measures the predictable sameness in a document.

MECHANISM SAMPLE

STYLE PORTRAIT

```
                  Sentence            Document
               1  2  3  4  5        1  2  3  4  5
Retention         X    ------             X ------
Verb Strength    -X----------        -X-----------
Formal        ------------        ------------
Verb Clarity      X ---------      X    ---------
Action        -------------X-     ---------------
Unclear       -X----              ------
Sensitivity   ----X-              ------
Climactic     ----X----           ---------
Theme Level    X       ------         X    ------
Incorrect
Filler        ---                 ---
Slang
Reflection    ------              ------
Complexity    ----X-              -X----
Ambiguity     -X-                 ---
Visual Form   ---------           ---------
Literary      ---                 ---
Derogatory
Breezy
Abstract      ----X-              -X----
Difficulty    ---                 ---
Monotony      ------ X            -X----
```

```
-----   preferred setting area
  X     actual setting
'Sentence' shows the last sentence's statistics.
'Document' shows the cumulative document average.
```

Style	Base	Explanation
Retention	5	How well your reader will remember what was written.
Verb Strgth	5	Measures your choice of good, strong verbs.
Formal	1	Measures how formal your writing is.
Verb Clarity	5	Measures how well actions are being expressed.
Action	5	Measures the verb strength of the main clause.
Unclear	1	Measures the use of unclear words or phrases.
Sensitivity	1	Measures the potential level of mis-communication.
Climactic	1	Is your most important information last?
Theme Level	5	Measures the use of strong topics, actions, focuses.
Incorrect	1	The number of incorrect words or phrases used.
Filler	1	Measures the number of times filler words are used.
Slang	1	Measures the amount of slang words in a sentence.
Reflection	1	Measures the level of author intrusion.
Complexity	1	Measures the complexity of the sentences.
Ambiguity	1	Measures the potential ambiguity in the document.
Visual Form	1	Abbreviations, foreign expressions...
Literary	1	Measures the use of literary words and phrases.
Derogatory	1	Measures the use of derogatory words and phrases.
Breezy	1	Measures the use of folksy and informal expressions.
Abstract	1	Measures the use of abstract nouns in the document.
Difficulty	1	Measures the use of jargon, legalese, code words...
Monotony	1	Measures the predictable sameness in a document.

Distribution List

Distribution:
Office of Naval Research (Code 461), (Code 10)
Defense Technical Information Center (DTIC) (4)

Copy to:
Director, Acquisition Education, Training and Career Development Policy, OUSDA(A)
Assistant for Manpower, Personnel, and Training (N911H)
NAVAIRWARCEN TR DIV, Technical Library (5)
Chief of Naval Education and Training (Code 02)
Commanding Officer, Naval Education and Training Program Management Support Activity (Code 03) (2), (Code 04) (2)
Chief of Naval Personnel (PERS-6E)
Department of the Navy, Total Quality Leadership Office